

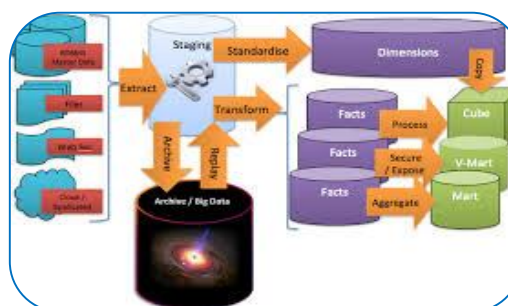


## DATA WAREHOUSE MODELLING AND QUALITY ISSUES AND ITS APPLICATIONS

**Jayashrei Shivajirao Kolthe**  
Research Scholar

### ABSTRACT :

A data warehouse is a database designed to enable business intelligence activity: it exists to help users understand and enhance the performance of their organization. It is designed for query and analysis rather than for the transaction process, and usually contains historical data derived from transaction data, but may include data from other sources. Data warehouse transactions separate the analysis workload from the workload and enable the organization to consolidate data from multiple sources.



**KEYWORDS :** data warehouse , transaction process , analysis rather.

### INTRODUCTION

Data Warehouse users perform data analysis that is often time-related. Examples include the consolidation of sales figures from the previous year, analysis of inventory and profitability by product and customer. But time-intensive or not, users want their data to be “sliced and diced”, and a properly designed data warehouse will be flexible to meet those demands. Users sometimes need highly aggregated data and other times need to drill down for details. More sophisticated analytics include trend analytic policies and data mining that use existing data to predict trends or predict futures. The data warehouse acts as a built-in engine used by the middleware business intelligence environment to provide end-to-end reports, dashboards, and other interfaces to users. Although the discussion above focuses on the term “data warehouse”, there are two other important terms that need to be mentioned. These are Data Mart and Operation Data Store (ODS). Data Mart plays the same role as a data warehouse, but is deliberately limited in scope. It can serve a specific department or business. The advantage of Data Mart against data warehouses is that it can be built much faster because of its limited coverage. However, data marts also cause problems with inconsistencies. It takes strict discipline to keep the data and computation definition consistent in Data Mart. This problem is widely known, so the data style exists in two styles. Separate data marts are those that are passed directly from source data. They can turn into islands of inconsistent information. Dependent data marts are provided from the existing data warehouse. Dependent data marts can prevent anomaly issues, but they must already exist for an enterprise-level data warehouse.

Operational data stores exist to support day-to-day operations. ODS data has been cleared and validated, but not historically deep: this may be just the data of the present day. Instead of supporting historically rich queries that the data warehouse can handle, ODS offers data warehouses a place to access even the most current data, which is not yet loaded into the data warehouse. ODS can also be used as a source for loading data warehouse. As data warehousing loading techniques become more advanced, data

warehouses may need less of an ODS as a source for loading data. Instead, static trickle-feed systems can load data warehouses in the near future.

In the data warehouse, data from multiple sources is brought into one place and then translated into a form that the data warehouse can process and store. For example, a business stores information about its customers, products, employees and their salary, sales and receipts. The boss may ask for the latest cost-cutting measures and analysis of all the data mentioned above will be necessary to get the answers. Unlike basic operational data storage, the data warehouse contains aggregate historical data.

#### APPLICATION OF DATA WARE HOUSE:

The data warehouse has deep-rooted applications in every industry that use structured and obsolete data from separate sources for forecasting, analytical reporting and business intelligence, allowing for tough decision making. Data warehouses are deeply rooted applications in every industry due to their potential that use historical data for forecasting, statistical analysis and decision making. Listed below are the applications of data warehouse against numerous industry backgrounds. Here are some of the major applications of data warehouses in different industries:

1. **Baking and Financial Industry:** In the banking industry, the focus is on the evolution of risk management and policy, as well as the customer data, market trends, government regulations and reporting, and more importantly, financial decision making. Many banks use warehouses to effectively manage the resources available on deck. Specific banking areas are used to conduct market research, analyze the performance of each product, exchange and exchange rates and develop marketing programs. Cardholder transaction analysis, cost patterns and merchant classification all have the opportunity to present special offers and benefit deals based on the cardholder's activities. In addition to all this, co-branding has potential. Now that we know some of the benefits of a data warehouse, let's look at how financial institutions use them together for profit.
2. **Goods Industry of Consumer:** Retailers and consumer-packaged goods (CPG) companies have a great deal of transaction data available: Every day, companies track every SKU sold to each customer at each store. In addition, companies regularly use state-of-the-art market-research techniques to answer a variety of questions: What products should you develop and sell? How much is the customer willing to pay? Which products should we discount and when? Which marketing vehicles will allow us to reach the most customers? The combination of Big Data and advanced analytics gives retail and CPG companies countless opportunities across the value chain. In portfolio strategy and product development, for example, companies can gain more detailed knowledge of customer needs and perspectives, and more accurately identify customer segments, which improves their ability to target high-value opportunities. They can measure return on investment (ROI) costs for marketing both traditional and new marketing vehicles (such as social media), allowing them to move marketing dollars to the most effective channels (see sidebar, "Exploring the Web: New Ways to Understand Online Customers"). Retailers can reduce out-of-stock by providing comprehensive hourly analysis of stock-by-rate rates by the store, giving customers a better buying experience and boosting sales for both themselves and their CPG partners.
3. **Government:** The government has a similar division between the two major subdivisions of the DWH industry. Governments can use data warehousing techniques in different fields e.g. Search for extremity profiles and risk assessments, in agriculture, education industry, finance department, medical department and fraud. There are many issues related to user fraud in the telecommunications industry and banking industry. General Information Services Terminal of the National Information Centre (GISTIC) Data Warehouse for Tamil Nadu was implemented in 1998. In this data warehouse, information on levels such as education, health, rainfall, poverty line data (BPL) was collected. Survey) etc. This is the Justnik Data Warehouse web enabled data warehouse that provides information on national issues on the latest issues in food, agriculture and

science and technology on a variety of topics. The World Bank collects and monitors huge data of the economic and development of all countries of the world. The World Bank began collecting and analysing data for the purpose of overseeing various World Bank projects in all countries of the world. For over 100 developing countries, the World Bank also captured their economics and financial credit data. The Ministry of Rural Development (MRD) is regularly reviewing data on rural. This large amount of data becomes difficult to analyse. But data warehouse technology makes analysis easy in that all analysis variables are taken into account. Rural development data warehouse can be developed by collecting more information about the district's infrastructure. Governments are playing a role in detecting any threats and fraud caused by unfortunate people. Unfortunately, almost no specific information warehouse implementation is available. Governments have access to a data warehouse, but they need a data warehouse system that is connected to every corner so that threats and terrorists are monitored.

4. **Education:** DWH is becoming more and more popular in the education sector. The use of DWH in academia presents several potential benefits for making appropriate decisions and evaluating data over time, which is a fundamental goal of the DWH process. DWH provides an integrated and holistic view of the organization. Related departments often use the data warehouse as a source of information about teachers and students. DWH helps students to take their results and notes from the web portable database instantly from the student portal and, last but not least, provide current and historical information about the organization. On a larger scale, DWH can integrate information from different organizations into a central repository for analysis and strategic decision making. In today's competitive environment where education is privatized and throat competition is accelerated, educational institutions need to be more organized and make better decisions. Educational institutions can overcome these kinds of problems by using the data warehouse. Not only did they set up operational data from complex recovery, there are many other benefits to management.
5. **Healthcare Environment:** Data warehousing has found no easy and easy way for medicine and healthcare. Data warehousing came up when retailers, financial institutions, government agencies, manufacturers and many other industries were collecting corporate information, healthcare and pharmaceutical bonuses. Data warehousing was still needed in health care and medicine, but the information requirements for drug and health care services were fundamentally different from other organizations. The first difference between the need for information about health care and medicine and other institutions was in the form of information. As an example of a fundamental difference against information found in a health institution, consider the facts that led to the gathering of information, as found in other institutions. As an example of the way corporations collect information, consider an airline. There are many transactions in airlines commerce. Reservation is provided. Flight departs. The goods are loaded. The seat is selected. Goods are delivered, and so forth. Practically every interaction between the customer and the airline has an impact on the transaction. Information is collected once the transaction is executed. The healthcare industry is evolving rapidly and providers and payers are finding that data tics analytics is becoming an essential tool. As a result of the increasing use of mobile medical devices and electronic records and healthcare CRMs, insurers and health care providers are collecting more and more data from large hospitals to small local hospitals. What to do with this data has become a hot topic, and health care providers are searching for what many healthcare payers already know: Data analytics is an essential tool.
6. **Hospitality Industry:** The hospitality industry, from hotels to cruise line operators and theme parks, can gain significantly deeper insights about their customers through the right applications of Big Data and hospitality management system software. More than other service industries, hospitality companies see a direct connection between customer experience and the satisfaction of their travel, stay, or trip. Understanding the factors that make interaction positive for customers, if combined

and split into smaller demographic groups, allows hospitality organizations to improve the performance of existing programs in an organized, effective and productive manner and determine the effectiveness of existing programs. General travellers plan or vacation their vacation in search of information aimed at using tourism products such as accommodation, tourism and entertainment. In addition, they seek information about other travellers' opinions, intending to visit them to find out if they have a good experience at destinations. Therefore, it is common for travellers to make their decisions using what they want to use in the Holy Days and access information on the Internet. In the resume, they will buy according to their taste and opinion of other travellers. The amount of information related to tourism activities is too large to represent all the different activities related to the tourism sector in tourism and hospitality. When tourists plan their trip, they seek opinions about aspects of their choice and similar sights made by other travellers. When a destination is selected and purchased, they have been purchasing the product based on the information available on the web. Relevant tourism information that exists on the Internet is constantly updated and has to be managed multiple times a day to reflect business needs and customer needs.

7. **Insurance Industry:** Insurance can be the first data-analyst business to evolve long before the age of automation. The benefit of the insurance industry requires the ability to calculate high risk and reward with high risk. In the economic research of real scientists, both the descriptive and predictive fields of statistics may have done much to move beyond the combined stimulation of the natural sciences. Yet this extreme intensity of the analysis of aggregate statistics created a blind spot for data analytics. The industry specializes in basic or expected behaviour, but misses the sharp differences that predict different results. And surprisingly, the big picture about the line, the market, or the benefits to consumers is unclear. Late in the decade, insurance companies began using data warehousing technology to create new business opportunities. They did this by exposing new perspectives at both the macro and micro levels. The macro view allows them to understand the market or the overall geography by integrating a wide array of external sources, such as demographic, psychographic, public records and locator data, with an internal experience base. Micro View uses pattern recognition techniques to reveal hidden, and sometimes counter-intuitive, facts and associations used to create new products, services, or processes thereafter.
8. **Telecommunication Industry:** There is a wealth of opportunity for the telecommunications industry to accept the challenge of providing data warehousing capabilities, but data collection and analytical requirements may push the limits of current technology. Some markets are moving faster than telecommunications. Some promising products - and in fact, most companies - did not exist more than a year ago. The basic identity of this market has changed profoundly here. The three forces combined to operate telecommunications in the most competitive business sectors of today's market: technology development, user demand and blockchain. Ten years ago, the heart of the telecommunications industry was long-distance service. Now it is technically possible for new services from paging to cellular phones and wireless internet. Users are signing it and demanding ever-increasing quality (not to mention bandwidth). The blockade also allowed competition to open in traditional areas such as local phone and cable services. It has also led to a series of mergers and acquisitions in both small companies and industry giants. All that leads to one bottom line: the data management nightmare. Large, long-established companies have legacy data on mainframes that may need to be integrated with data stored somewhere on a series of new products. Worse, data from a small company purchased to deliver a new product may need to be combined with that legacy data. Most new, smaller companies have easier integration tasks, but there are limitations in resources. Yet the command to "know the customer" means that no matter how challenging, the task must be achieved or the company will not survive.

---

**CONCLUSION:**

Simple data profiling can be used to detect some data quality issues in the data. There are many types of data quality analysis for different types of data content. This always requires knowledge about the business that is used to create the right data quality rules. Interviews were conducted in the data warehouse to gain knowledge about what types of data quality clients and suppliers face. Many common data quality problems input data incorrectly, and the second most common problem is missing or incomplete data.

**REFERENCES:**

1. Aravind Kumar (2016), 'Exploring the Application of Data Warehouse in Public Life', International Journal of Computer Science and Information Technologies, ISSN – 0975-9646, Vol-7, Issue-6, pp. 2544-2553
2. Hensen H., Kremar H., and Scheman M. (2014), 'Big Data - An Interdisciplinary Opportunity for Information Systems Research', Business and Information System Engineering, Vol-6, Issue-5, pp. 261-266.
3. Manjunath T. N., Ravindra S. Hegadi, Umesh I. M., and Ravi Kumar G. K. (2012), "Realistic Analysis of Data Warehousing and Data Mining Application in Education Domain" in International Journal of Machine Learning and Computing, Vol. 2, No. 4
4. Pinar, B. (2005), "A Comparison of Data Warehouse Design Models." PhD diss., Master Thesis, Atılım University
5. Romero O., and Abello A. (2009), 'A Survey of Multidimensional Modeling Methodologies', International Journal of Data Warehouse Min, Vol-5, pp. 1-23
6. Mukherjee S. (2018), 'Perception and conceptualization of Data Warehouse of a day to day Transaction', International Journal of Engineering and Computer Science, ISSN 2319-7242, pp. 24130-24151